# Robust Range Image Registration using a Common Plane

Joachim Bauer
VRVis Research Center
Inffeldgasse 16
A–8010 Graz
Austria
bauer@icg.vrvis.at

Konrad Karner
VRVis Research Center
Inffeldgasse 16
A–8010 Graz
Austria
karner@vrvis.at

Andreas Klaus
VRVis Research Center
Inffeldgasse 16
A–8010 Graz
Austria
klaus@vrvis.at

Roland Perko
University of
Technology Graz
Inffeldgasse 16
A–8010 Graz
Austria
perko@icg.tu-graz.ac.at

## ABSTRACT

This paper presents a method for the registration of range image pairs that were captured from different locations. The proposed approach assumes that a common 3D plane is least visible in both scans. Based on this assumption the number of unknown transform parameters reduces from six to three. An orthogonal rectification process is then used to convert the 3D data into 2D depth images and correspondences that are found by a simple 2D image matching process are used to determine the remaining unknowns. We propose a robust plane detection algorithm to detect 3D planes in the scans and describe the algorithm for the computation of the ortho-rectified range images. A correlation based image matching process is used to establish correspondences between the rectified depth images. The experiments are carried out on typical facade data from a city scanning project.

## Keywords

Range image, Registration, Matching

## 1 Introduction

Using laser scanning devices for the fast acquisition of 3D objects has become a popular topic in the last years. The goal of the digital Michelanglo project [Levoy00] was the development of techniques for capturing, post-processing and visualization of sculptures and architecture of Michelangelo. Usually a general 3D object can not be modeled from a single scan - this is due to occlusion from other objects, self-occlusions i.e. concavities in the objects and back-facing parts of the object. Generally an object is therefore scanned from several viewpoints. A fundamental step in the modeling pipeline is thus the registration of the separate scans into a common coordinate system. A similarity transform with six degrees of freedom relates 3D points from different scans. For outdoor applications such as city modeling or the scanning of other large objects the position and orientation of the scanner is unknown and must be either provided by other instruments e.g. a total station or extracted from the point clouds. For the robust detection of this inter-scan transform two subsequent scans must at least partially cover an identical part of the scanned object. The points from this overlapping region between two scans can then be used to compute the similarity transform that optimally aligns the fraction of points in the scans that specify the same physical part of the scanned object. The most popular approach for the registration of overlapping point clouds is the Iterative Closest Point Approach (ICP) developed by Besl and McKay [Besl92] and Chen and Medioni [Chen92]. The ICP algorithm is the basis of almost all registration algorithms.It basically iteratively pairs each point in the first set to its closest neighbor in the second set and computes the transform that minimizes the error between the two scans. Since this approach was very successful many different variants of the original algorithms have been proposed. The ICP method is however limited to fine registration of point clouds - that means that an coarse alignment must be already given. In the case of scanned facades where many self similarities like windows, balconies, doors etc.


***WSCG POSTERS proceedings***
*WSCG'2004, February 2–6, 2004, Plzen, Czech Republic.*

occur, a purely ICP-based approach often fails or converges to a local minimum. In the following section we propose an approach for the coarse alignment of range image pairs and the subsequent sections describe the different steps that are necessary to compute the similarity transform. Experiments with real range data and some remarks on future work conclude the paper.

## 2 Our approach

We propose a method for the coarse alignment of 3D point clouds using extracted 3D planes. As stated by Eggert et al. [Egger98] the registration of point clouds can be achieved easily if 3 corresponding non-parallel planes are given in each scan - two corresponding planes in each scan solve the alignment problem up to one translational degree of freedom. In the case of city scanning, when the acquisition platform moves along a street with nearly parallel buildings, it is often not possible to extract more than the main facade plane (the ground plane is often occluded by parking cars, vegetation etc.). Our approach assumes that in both scans at common 3D plane is at least partially visible. In this case the number of unknowns drops from six to three, leaving two translational parameters and one rotation angle unknown. We use the three transform parameters that are already determined by the common 3D plane to transform the point sets into the x-y plane. After this transform we can view the scans as a 2.5D height-field and compute a orthogonal rectified 2D depth image from the 3D point set. These orthogonally rectified (ortho-rectified) range images have the advantage, that all features in the scanned object are invariant against rotation and translation. Figure 1 shows the differences between the original range image and the ortho-rectified range images. Whereas the two objects (circular and quadratic) occluding the plane are mapped to different positions in the original range images (the view frustum of each scan is illustrated as a triangle), they are on the same position in the ortho-rectified range image (illustrated by the exemplary dark gray profiles). In this ortho-rectified depth images we search for corresponding points and use the correspondences to set up rigid 2D transform that incorporates the remaining three unknowns.The work-flow we propose can be divided into the following steps: First we extract 3D planes from the two given scans (using the method described in section 3) and compute the ortho-rectified range images (see section 4). For these ortho-rectified range images we try to find corresponding points in by an image matching step (see section 5) and finally calculate the remaining unknowns from the point correspondences and compose the final 3D similarity transform for the point clouds. The final similarity transform $S$ is now composed from the transform $T_1$ that is determined by the 3D plane and the transform $T_2$ that is computed from the point correspondences that are extracted by the image matching step.

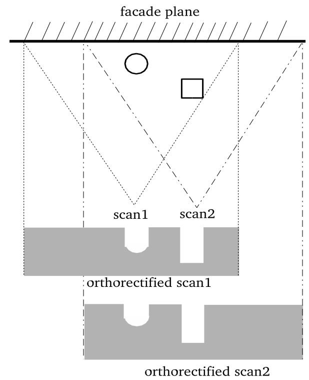

Figure 1: Visualization of the orthogonal rectification. The top view of a facade plane with two occluding objects (circular and quadratic) in front. In the original view frustum of the two scans the occluding objects are mapped on different positions in the range images. The exemplary profiles of the ortho-rectified range images show, that the position of the objects relative to the plane is now invariant to the view point of the scanner.

## 3 Plane extraction

The segmentation of 3D point clouds has been addressed in several publications - a comparative study is given in [Hoove96]. Eggert et al. [Egger98] presented a method for the segmentation of range images of industrial parts. The geometry of buildings typically consists of several planar faces - the main facade planes. In densely packed areas however only one or two of the facade planes are visible. Another problem is that adjacent planes are not necessarily parallel e.g. due to a slightly curved street. Nearly parallel planes present a challenging situation for the automatic plane extraction. We propose an approach based on the RANSAC principle [Fisch81] and perform it on a number of local subsets of the original point cloud. The following iterative method is employed to detect 3D planes in the 3D point cloud: First we randomly choose a set of seed points from the point cloud (the number of seed points depends on the number of planes present in the point cloud) and get the subset of points lying in a local neighborhood of each seed point. Subsequently a RANSAC based plane detection algorithm is applied on these local point sets around the seed points and the resulting local plane hypotheses are verified, discarding multiple identical solutions. A final refinement of the remaining local RANSAC solutions with an iterative least squares fit using the whole point cloud yields robust plane hypotheses. For fast spatial indexing the points of one scan are stored in a 3D KD-tree. The iterative refinement of the detected RANSAC plane hypothesis uses only those inlier points that form a densely connected set - this is comparable to a region growing in 2D. The largest connected set is found by recursively searching the inlier point cloud from different seed points. This approach minimizes the influence of points that belong to other closely adjacent planes or coplanar but distant planes. Finally the planes that share too many mutual inliers are rejected. The detection method is comparable to the approach of Danuser and Stricker [Danus98] and yields a set of accurate 3D planes.

## 4 Orthogonal Rectification

The process of ortho-rectification is used to transform the original range image, which is a central perspective image, into a parallel projection image. In order to be able to perform a projection along a principal axis we have to transform the original point cloud of the scan such that the normal vector of the extracted plane points into the z-direction of the coordinate system. This transform, which is a simple 3×3 rotation matrix $R$, allows us to view the scan points as a 2.5D heightfield (given that for a specific x-y pair a unique z value exists). The rotation matrix is set up using the normal vector $n$ of the extracted 3D plane and two orthogonal side vectors $s_1$ and $s_2$. After applying the transform the normal vector $n$ is parallel to the z-axis and the side vectors are be parallel to the x-axis and the y-axis respectively. The rotation matrix is given as: $R = [s_1, s_2, n]$. We also apply a translation $t$ to shift the rotated points so that the 3D plane lies in the x-y plane. The translation vector $t$ is: $t = -Rp$, where $p$ is a point on the 3D plane. Hence the 4×4 transform matrix $T_1$ is given as: $T_1 = [R|t]$. For the efficient computation of the height values for each pixel of the orthogonally rectified range image we triangulate the point set and store the triangles in an octree. The octree is, like the KD-tree for 3D points, a data structure for the efficient spatial access of triangles and other geometric objects. If the scanner system provides a range image the triangulation is inherently given - for an irregular point cloud a standard Delaunay triangulation in 2D is performed. Triangles with very long sides (compared to the scanner step width) indicate occlusions or non valid measurements and are eliminated from the triangulation. With this approach it is also possible to restrict the depth range of the resulting range image. The pixel values of the orthogonally rectified range image are computed by intersecting the triangle mesh with rays that are parallel to the normal vector (which is $(0, 0, 1)$ after applying the transform $T_1$). In case of a holes (gaps) in the mesh, where no valid intersection is found, a predefined non valid depth value is stored.

(a) (b)

Figure 2: Orthogonal rectification: (a) original range image with severe distortion. (b) ortho-rectified range image with nonlinearly coded depth values.

For the subsequent image matching process we must map the depth range of the scan to the intensity range of the output image. In the case of facades a lot of information is contained in the structure of the facade: examples are indentations of windows and doors, protrusions of friezes and balconies etc. On the other hand we also want to map objects that are farther away from the facade, since they provide valuable information for the subsequent matching process (especially if repeating patterns like identical windows occur in the facade plane). In order to incorporate all this information in the final rectified range image we relate all depth values to the detected main facade plane and apply a nonlinear encoding of the depth values. With this nonlinear encoding we achieve a high resolution for structures that are near the plane and a reduced resolution for objects that are far from the plane. The function used for this nonlinear coding is the sigmoid function: $y = \frac{1}{1+e^{-x}}$. We shift the function to be symmetrical around the x-axis and use the following equation to compute 8 bit intensity values from the original depth values. Figure 2(a) shows the original range image and Figure 2(b) the ortho-rectified range image of a facade plane. Due to the nonlinear encoding of the depth values the micro-structure in the facade plane as well as object distant from the plane are preserved in the 8 bit depth image.

## 5 Matching

In the image matching step we compute a set of corresponding points between two ortho-rectified range images in order to set up a transform for the remaining three unknowns (the translational component $(t_x, t_y)^T$ and the in-plane rotation angle $\varphi$). Occluding objects that prevented the robust detection of the ground plane give now valuable clues for the matching process. We use the Harris corner detector [Harri88] to detect points of interest, normalized cross correlation and a least squares method for refining the results to subpixel accuracy. From the point correspondences we set up a system of equations to compute the unknowns using least squares adjustment: The final similarity transform is the combination of $T_1$ and $T_2$: $S = T_2 T_1$.

## 6 Results

The range images used in this examples were captured with the Riegl LMS Z210 and the Riegl LMS Z360 laser scanner and have a resolution of 424×754 pixels. The accuracy of the measured points $\pm 2.5cm$. Figures 3(a) and (b) show a pair of orthogonally rectified range images and figure 3(c) shows the registration result for the range image pair - the corresponding points are denoted as black crosses. The overlap of the scans is more than 50 percent. The average 3D registration error computed on 50 percent of the best matching points is 4.11cm. After performing an ICP registration the error dropped to 1.87cm, which is lower than the measurement uncertainty. This experiment shows, that a coarse registration based on our proposed method is possible and that a subsequently applied ICP registration successfully converges to a global minimum.

## 7 Conclusion

We presented a method for the coarse alignment of range image pairs under the condition that both scans map the same 3D plane. A robust method for the detection of 3D planes based on the RANSAC principle was introduced. An orthorectification process, using the extracted 3D plane, was used to produce range images without perspective distortions. A correlation based matching approach establishes point correspondences between the two range images and these correspondences are used to set up a transform that solves the three remaining parameters of the similarity transform that aligns the 3D points of the scans. Further improvement in the matching accuracy can be achieved by integrating the reflectance information of the scanned points - this is standard feature of modern laser scanners such as the Riegl scanners. The reflectance value can be seen as texture information for the measured 3D points.

## Acknowledgments

This work has been done in the VRVis research center, Graz and Vienna/Austria (www.vrvis.at), which is partly funded by the Austrian government research program Kplus.

## References

[Besl92] P. J. Besl and N.D. McKay. A method for registration of 3d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–259, 1992.

[Chen92] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10:145–155, 1992.

[Danus98] G. Danuser and M. Stricker. Parametric model fitting: From inlier characterization to outlier detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 1998.

[Egger98] D.W. Eggert, A.W. Fitzgibbon, and R.B. Fisher. Simultaneous registration of multiple range views for use in reverse engineering of CAD models. *Computer Vision and Image Understanding*, 69(3):253–272, 1998.

[Fisch81] M.A. Fischler and R.C. Bolles. RANSAC random sampling concensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of ACM*, 26:381–395, 1981.

[Harri88] C. Harris and M. Stephens. A combined corner and edge detector. *Proceedings 4th Alvey Visual Conference*, 1988.

[Hoove96] A. Hoover, J.-B. Gillian, X. Jiang, P.J. Flynn, H. Bunke, D. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, and R.B. Fisher. An experimental comparison of range image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7):673–689, 1996.

[Levoy00] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital Michelangelo project: 3D scanning of large statues. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 131–144. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[Rusin01] S. Rusinkiewicz and M. Levoy. Efficient

(a) (b) (c)

Figure 3: Orthogonal rectified range image pair and registration result: (a) right image. (b) left image (c)Registration result for the range image pair. The corresponding points are shown as black crosses.

variants of the ICP algorithm. *Proc. Third Intern. Conf. on 3D Digital Imaging and Modeling*, pages 145–152, 2001.